Exhibit A

Disclosure YOR8-2003-0463

Prepared for and/or by an IBM Attorney - IBM Confidential

Created By Andy Aaron On 06/20/2003 04:48:47 PM EDT
Last Modified By Emily Forbes On 01/07/2004 09:14:27 AM EST

Required fields are marked with the asterisk (*) and must be filled in to complete the form .

*Title of disclosure (in English)
Method of Improving TTS Intelligibility in Long Passages

Summary

Status	Final Decision (File)
Final deadline	
Final deadline reason	
Docket family	YOR9-2004-0004
* Processing location	Yorktown
* Functional area	(907) 907 Human Language Technologies (speech, gesture, ink,...)
Attorney/Patent professional	Thu A Dang/Watson/IBM
IDT team	David Nahamoo/Watson/IBM Thu A Dang/Watson/IBM Michael Picheny/Watson/IBM
Submitted date	07/16/2003 10:02:22 AM EDT
* Owning division	RES
Incentive program	
Lab	
* Technology code	653
PVT score	

Inventors with a Blue Pages entry

Inventors: Andy Aaron/Watson/IBM, Ellen Eide/Watson/IBM

Inventor Name	Inventor Serial	Div/Dept	Inventor Phone	Manager Name
> Aaron, Andrew	4A4931	2C/GVCA	862-2078	Sakrajda, Andrzej (Andy)
Eide, Ellen M.	707218	22/S0IA	862-1177	Picheny, Michael

> denotes primary contact

Inventors without a Blue Pages entry

IDT Selection

Attorney/Patent professional	Thu A Dang/Watson/IBM
IDT team	David Nahamoo/Watson/IBM Thu A Dang/Watson/IBM Michael Picheny/Watson/IBM
Response due to IP&L	08/17/2003

Main Idea

To view the main idea for this disclosure, click on this doclink --->  (If you are prompted to enter a server name, please enter D01DB068/01/A/IBM)

Evaluation

This team evaluation was entered by Michael Picheny/Watson/IBM on 10/13/2003

What is the team's evaluation of this disclosure? Search

Date evaluated : 10/13/2003

Evaluation comments

Final Evaluation History	Who made the final evaluation	Final evaluation date

Search	Michael Picheny/Watson/IBM	10/13/2003

Search Information

Date sent:	*Target completion date:	Search results received date:
Who was the search sent to (This area is to designate a Local Searcher name or WAIPL):		
*Search type: ☐ Patentability ☐ Clearance ☐ Validity ☐ State of Art		
*Features to be searched:		

Search Office Information

Target completion date:	☐ Search has been delayed	Ship/Return date:
Search conducted by		

Comments

Final Decision

This decision was entered by Emily Forbes/Watson/IBM on 01/07/2004

Decision: File	Status: N/A
PPM area: 600 - Software/Services/Applications/Solutions	
Date of final decision : 01/07/2004	

Additional filing information
Planned Filing date:
Filing comments:

Additional decision comments

Final Decison History

Entered on 7-Jan-2004 by Emily Forbes
File N/A 7-Jan-2004 Docket Family: YOR920040004

Post Disclosure Text & Drawings

To add additional information related to this disclosure once it has been submitted, click the action button below and a new document will be opened for you to enter the new information. To view existing post disclosure information, double-click on the item in the list below (if there has been additional information entered), and the document will open for you to view.

Date entered	Post disclosure comments and drawings (double-click an item below to view)

Form Revised (05/28/03)

Main Idea for Disclosure YOR8-2003-0463
Prepared for and/or by an IBM Attorney - IBM Confidential
Archived On 10/14/2003 01:04:18 AM

Title of disclosure (in English)
Method of Improving TTS Intelligibility in Long Passages

Main Idea

1. Background: What is the problem solved by your invention? Describe known solutions to this problem (if any). What are the drawbacks of such known solutions, or why is an additional solution required? Cite any relevant technical documents or references.
Text-to-speech software ("TTS") has made vast improvements in the past 1-2 years. What used to be a serviceable but robotic-sounding system now mimics the human voice with great fidelity. But paradoxically, the increased fidelity leads to an increase in perceived faults -- as the sound gets closer to that of a live human, all of its shortcomings come more clearly into view.

And one of those limitations is its failure to hold the attention of a listener for long passages. While we plan to use TTS to play back news stories and long emails, its limited prosodic richness and monotonous tone present a barrier. When listening to a long passage, there are sections of great clarity and punctuated by occasional words or word groups that are harder to understand, or that suffer from bumpy synthesis. These junctures present an increased cognitive load, and the listener must work harder to decipher what he or she has just heard. Meanwhile, the TTS marches on, so while the listener is working out the previous word the software is busy producing new ones. The end result is listener fatigue. It feels as though the TTS is being insensitive to the needs of the listener, whose mind ultimately begins to wander.

There are no current solutions to this problem.

2. Summary of Invention: Briefly describe the core idea of your invention (saving the details for questions #3 below). Describe the advantage(s) of using your invention instead of the known solutions described above.
Let's look at a sentence from a news report:

> "'Bank of America tends to be a pretty good litmus test for the financial services sector as a whole,' said Doug Lister of Wachovia Securities, a financial services company."

Most of this text will sound quite good coming out of the TTS engine. But when we get to the unfamiliar name "Doug Lister," we're on shaky territory. Did the TTS engine just say "Doug Lister," or was it "Doug Glister"? We've never heard either name, they're equally likely, and would sound pretty much the same. And while we're pondering that, the engine is generating still more words, until it gets to "Wachovia." Now we have to decode another word that we're not so sure about. Was that "Wockovious Securities," or "Wock Ovia Securities"? No, it was "Wachovia Securities." And with enough of these incidents, we begin to feel as though we're working too hard and we fall behind, ultimately missing some vital content.

Live news readers compensate for this problem by slowing down slightly at unfamiliar words and by adding an imperceptible pause before and after the words. They may even allow themselves to sound slightly hesitant. This does two things -- it signals the listener to pay extra attention, and it gives the listener some time to catch up. A live news reader would therefore read. "'Bank of America tends to be a pretty good litmus test for the financial services sector as a whole.' said *Doug Lister* of *Wachovia Securities*, a financial services company."

While our current TTS systems don't truly "understand" the content of their speech to the point where we could program them to know what words to emphasize, some of these problems areas are in fact predictable and therefore lend themselves to software solutions.

3. Description: Describe how your invention works, and how it could be implemented, using text, diagrams and flow charts as appropriate.

What we're trying to do is to determine in advance which words or word pairs are likely to suffer from uneven synthesis. There are several metrics that can be employed. For example, the TTS System has a dictionary and can spot words that are not in it. The front end can recognize capitalization rules. Therefore, it can with some reliability detect unfamiliar proper names, which have a high likelihood of synthesis problems. So when one is detected, a very small pause can be added, and/or the word can be synthesized with longer durations.

We could also make use of a statistical language model trained on large amounts of text to spot low probability words and word sequences. Words or word sequences that receive a low probability score would be similarly treated with small pauses and longer durations.

Other metrics are available to the system to predict when a difficult word or word pair has been encountered. We can monitor the cost function during the synthesis process. Splices exhibiting high cost are likely to lead to uneven synthesis, and the same treatment can be applied.

False positives are no cause for concern. If the occasional well-synthesized word is played a little slower, this won't sound abnormal. But if we can catch a reasonable percentage of rough synthesis and treat it with kid gloves through the strategic application of pauses and duration control, we can greatly increase our overall comprehension.

What we're trying to do is to determine in advance which words or word pairs are likely to suffer from uneven synthesis. There are several metrics that can be employed. For example, the TTS System has a dictionary and so can spot words that are not in it. The front end can recognize capitalization rules. Therefore, it can with some reliability detect unfamiliar proper names, which have a high likelihood of synthesis problems. So when one is detected, a very small pause can be added, and/or the word can be synthesized with longer durations.

We could also make use of a statistical language model trained on large amounts of text to spot low probability words and word sequences. Words or word sequences that receive a low probability score would be similarly treated with small pauses and longer durations.

Other metrics are available to the system to predict when a difficult word or word pair has been encountered. If the segment matching falls below some quality threshold, we can assume this may lead to uneven synthesis, and the same treatment can be made available.

False positives are no cause for concern. If the occasional well-synthesized word is played a little slower, this won't sound abnormal. But if we can catch a reasonable percentage of rough synthesis and treat it with kid gloves through the strategic application of pauses and duration control, we can greatly increase our overall comprehension.

To summarize: When the TTS System encounters a section of low confidence or unknown words, it will add pauses and increase durations.

How text would be marked up by Rare Sequence Detection

Input Text	**Hello, Mrs. Wisniewski**
Normalized text	**Hello** P0 **missus wisnefsky**
Normalized text plus rare sequence detector	**Hello** P0 **missus** P1 <rare> **wisnefsky** </rare>

Andy Aaron
IBM Watson Research Center, Rm. 06-155
1101 Kitchawan Road
Route 134
Yorktown Heights, NY 10598
Phone 914-945-2078
Fax 914-945-4492

Ellen Eide
IBM Watson Research Center, Rm. 23-126C
1101 Kitchawan Road
Route 134
Yorktown Heights, NY 10598
Phone 914-945-1177
Fax 914-945-4490

Method of Improving TTS Intelligibility in Long Passages

By Andy Aaron and Ellen Eide

Text-to-speech software ("TTS") has made vast improvements in the past 1-2 years. What used to be a serviceable but robotic-sounding system now mimics the human voice with great fidelity. But paradoxically, the increased fidelity leads to an increase in perceived faults -- as the sound gets closer to that of a live human, all of its shortcomings come more clearly into view.

And one of those limitations is its failure to hold the attention of a listener for long passages. While we plan to use TTS to play back news stories and long emails, its limited prosodic richness and monotonous tone present a barrier. When listening to a long passage, there are sections of great clarity and punctuated by occasional words or word groups that are harder to understand, or that suffer from bumpy synthesis. These junctures present an increased cognitive load, and the listener must work harder to decipher what he or she has just heard. Meanwhile, the TTS marches on, so while the listener is working out the previous word the software is busy producing new ones. The end result is listener fatigue. It feels as though the TTS is being insensitive to the needs of the listener, whose mind ultimately begins to wander. There are no current solutions to this problem.

Let's look at a sentence from a news report:

> "'Bank of America tends to be a pretty good litmus test for the financial services sector as a whole,' said Doug Lister of Wachovia Securities, a financial services company."

Most of this text will sound quite good coming out of the TTS engine. But when we get to the unfamiliar name "Doug Lister," we're on shaky territory. Did the TTS engine just say "Doug Lister," or was it "Doug Glister"? We've never heard either name, they're equally likely, and would sound pretty much the same. And while we're pondering that, the engine is generating still more words, until it gets to "Wachovia." Now we have to decode another word that we're not so sure about. Was that "Wockovious Securities," or "Wock Ovia Securities"? No, it was "Wachovia Securities." And with enough of these incidents, we begin to feel as though we're working too hard and we fall behind, ultimately missing some vital content.

Live news readers compensate for this problem by slowing down slightly at unfamiliar words and by adding an imperceptible pause before and after the words. They may even allow themselves to sound slightly hesitant. This does two things -- it signals the listener to pay extra attention, and it gives the listener some time to catch up. A live news reader would therefore read, "'Bank of America tends to be a pretty good litmus test for the financial services sector as a whole,' said *Doug Lister* of *Wachovia Securities,* a financial services company."

While our current TTS systems don't truly "understand" the content of their speech to the point where we could program them to know what words to emphasize, some of these problems areas are in fact predictable and therefore lend themselves to software solutions.

How it works

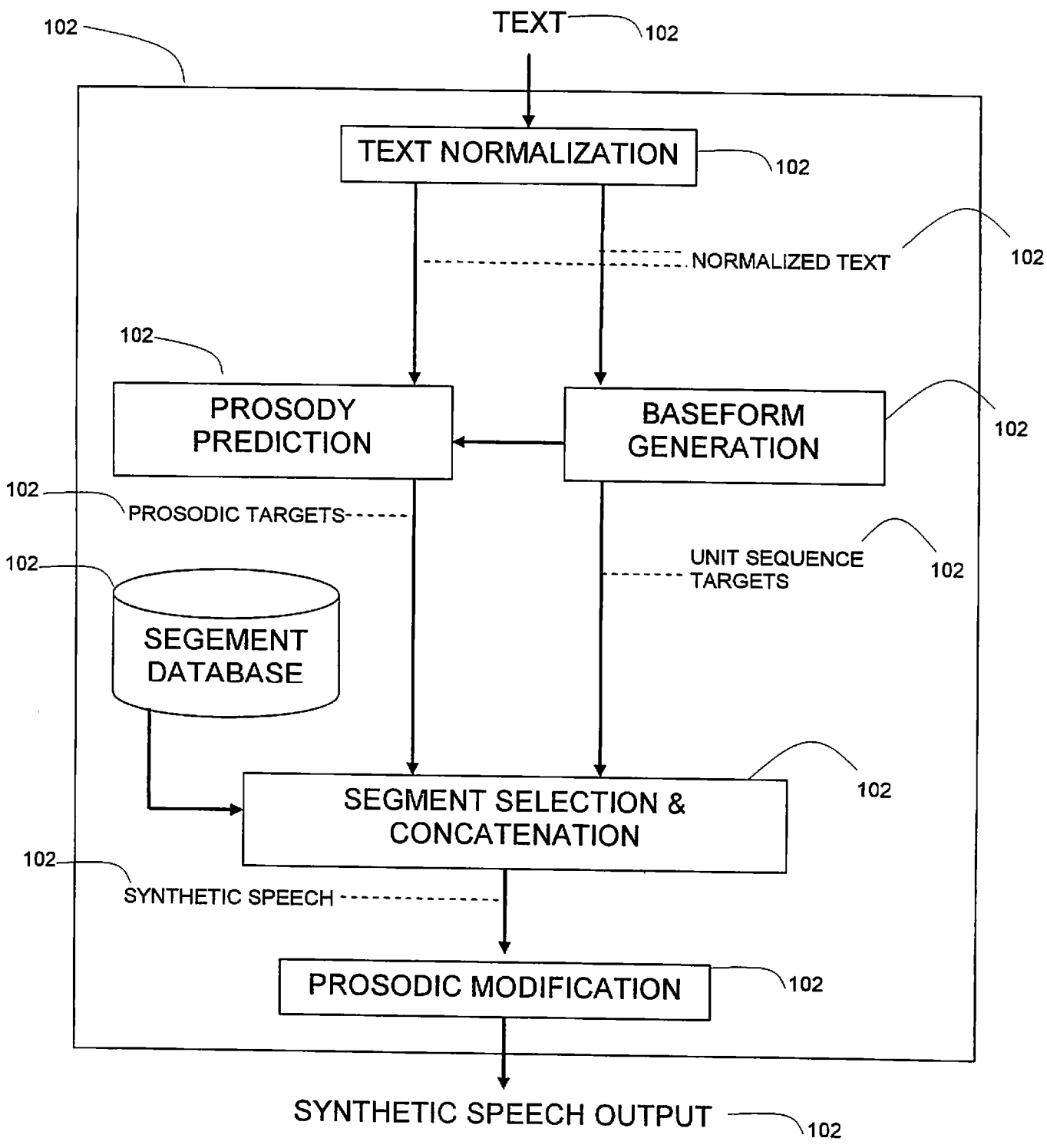
102
TEXT
102
TEXT NORMALIZATION
102
NORMALIZED TEXT
102
102
PROSODY PREDICTION
BASEFORM GENERATION
102
102
PROSODIC TARGETS
UNIT SEQUENCE TARGETS
102
102
SEGEMENT DATABASE
102
SEGMENT SELECTION & CONCATENATION
102
SYNTHETIC SPEECH
PROSODIC MODIFICATION
102
SYNTHETIC SPEECH OUTPUT
102

FIG. 1

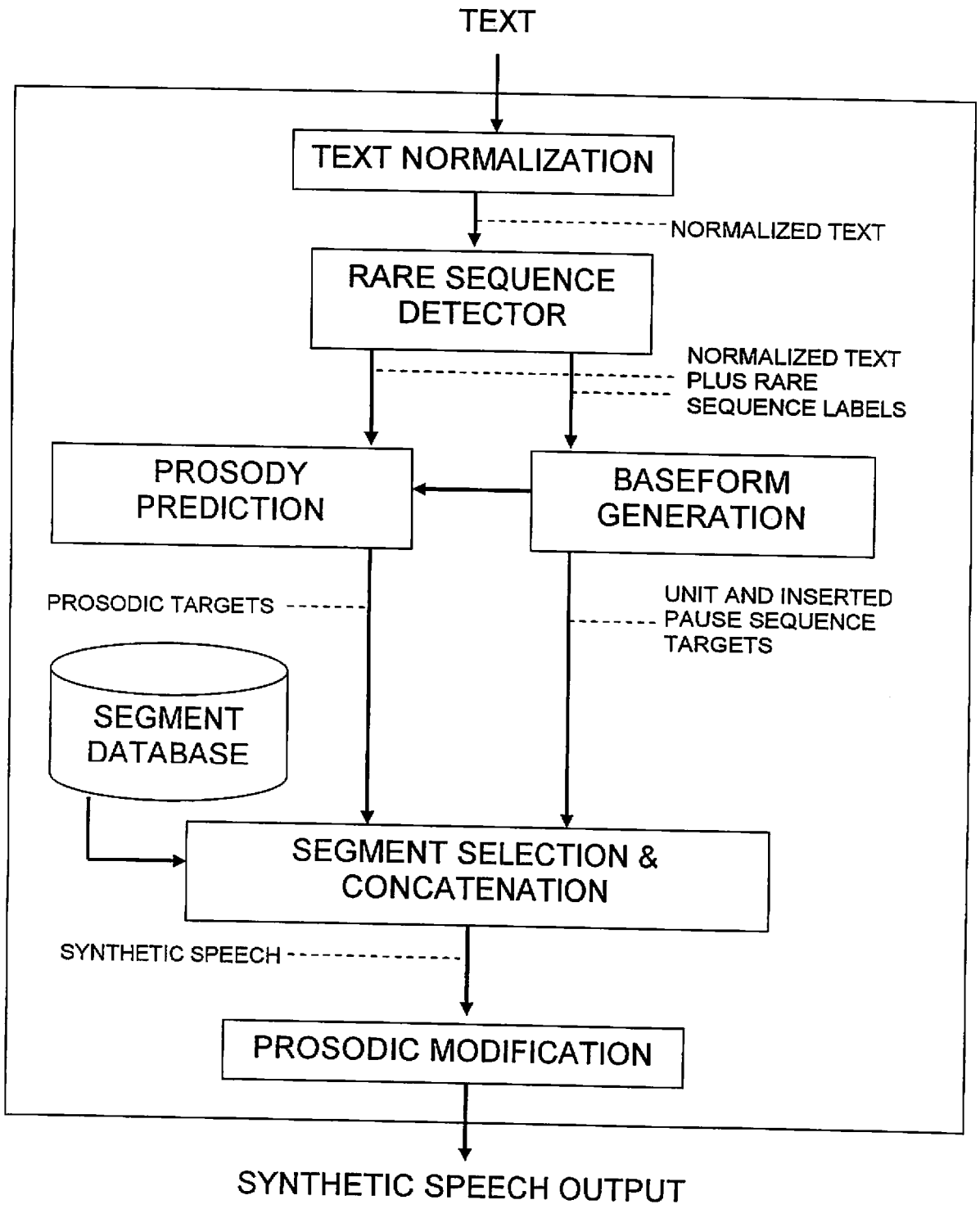
TEXT
TEXT NORMALIZATION
NORMALIZED TEXT
RARE SEQUENCE DETECTOR
NORMALIZED TEXT PLUS RARE SEQUENCE LABELS
PROSODY PREDICTION
BASEFORM GENERATION
PROSODIC TARGETS
UNIT AND INSERTED PAUSE SEQUENCE TARGETS
SEGMENT DATABASE
SEGMENT SELECTION & CONCATENATION
SYNTHETIC SPEECH
PROSODIC MODIFICATION
SYNTHETIC SPEECH OUTPUT

FIG. 2